Transmembrane Helix Assembly by Window Exchange Umbrella Sampling

Soohyung Park, Taehoon Kim, and Wonpil Im*

Department of Molecular Biosciences and Center for Bioinformatics, The University of Kansas, 2030 Becker Drive, Lawrence, Kansas 66047, USA

(Received 1 December 2011; published 8 March 2012)

A method of window exchange umbrella sampling molecular dynamics simulation is employed for transmembrane helix assembly. An analytical expression for the average acceptance probability between neighboring windows is derived and combined with the first passage time optimization method to predetermine a parameter set in an optimal range. With the parameter set, the method provides a substantially more efficient sampling of helix-helix interfaces together with the potential of mean force along the helix-helix distance of a transmembrane helix-dimer model, compared to the umbrella sampling method.

DOI: 10.1103/PhysRevLett.108.108102 PACS numbers: 87.10.−e

Membrane proteins are involved in many vital cellular processes [1]. Unlike globular proteins, the orientation of each helix in membrane proteins with respect to membrane bilayers defines their structures, which is closely related to their functions. Structures of various polytopic membrane proteins in detergents and micelles have been determined by crystallography and spectroscopy. However, it remains challenging to obtain structural information of membrane proteins with small numbers of transmembrane (TM) helices and their oligomers in bilayer environments [2]. Membrane proteins with a single-pass TM helix are abundant, and receptors with a single-pass TM helix make up about 30% of human TM receptors [1]. In this context, it is critical to have a reliable computational approach to provide structural models of these membrane proteins and the helix association energetics.

In the computational TM-assembly modeling, multiple degrees of freedom for helix motion, such as helix-helix distance, crossing angle, and rotation angle along each TM helical axis, need to be considered. Therefore, the modeling itself is computationally intensive even in finding the interfacial (contact) residues of a simple bitopic TM helix dimer [3]. Furthermore, in most molecular dynamics (MD) simulations, it usually has been assumed that the interface between mutated TM helices resembles that of the wild type, e.g., the right-handed interface of point mutated glycophorin *A* [4]. Without a prior knowledge of the mutant structure, such an assumption may be problematic because the configurational space sampling (e.g., crossing and rotation angles) at short or intermediate TM helix separations would be incomplete due to the strong interactions between the interfacial residues. Replica exchange (REX) methods [5–8] may overcome such difficulties by facilitating the sampling with the aid of regular exchange attempts between replicas at different temperatures. But these methods generally do not provide the potential of mean force (PMF) along the reaction coordinate(s), and their applicability to all-atom explicit lipid bilayers is limited, in which one can capture realistic helix-helix interactions. Recently, there has been remarkable progress in advanced sampling methods [9]. Among them, the Wang-Landau scheme [10] and the orthogonal space random walk (OSRW) strategy [11] are of particular interest. These methods, in principle, are quite general and can address all the aforementioned issues. Yet, these methods require sophisticated recursion procedures, and are therefore challenging to implement.

These issues for TM helix assembly can be addressed by window exchange umbrella sampling MD (WEUSMD), a variant of Hamiltonian REXMD [12], in which replicas (i.e., windows) along a certain helix motion (i.e., reaction coordinate) are exchanged. In particular, we derived an analytical expression to predetermine the parameter set in an optimal range, based on the first passage time optimization for REXMD [13], to make WEUSMD efficient. Thus, its application is straightforward compared to the Wang-Landau and OSRW methods. In the following, we first describe the analytic expression of the average acceptance probability between neighboring windows to determine a parameter set. Then, a practical application of WEUSMD with an optimized parameter set to a TM helix-dimer model is presented and discussed.

Let us consider a system consisting of N windows along a reaction coordinate ξ. The potential energy of the ith window with a coordinate set $\mathbf{R}_i$ under the window potential $w_m(\xi_i)$ is given by $U_i(\mathbf{R}_i) = U_0(\mathbf{R}_i) + w_m(\xi_i)$, where U_0 is the unbiased potential energy, $w_m(\xi_i) = k_m(\xi_i - \xi_m)^2/2$, $\xi_i \equiv \xi(\mathbf{R}_i)$, k_m is the window force constant, and ξ_m is the target value of the reaction coordinate. Assuming that the temperature (T) is constant for all the windows, the canonical distribution of ξ_i becomes

$$p_m(\xi_i) = \frac{\exp\{-\beta[\mathcal{W}(\xi_i) + w_m(\xi_i)]\}}{\int d\xi_i \exp\{-\beta[\mathcal{W}(\xi_i) + w_m(\xi_i)]\}}, \quad (1)$$

where $\beta \equiv 1/(k_B T)$ is the inverse temperature, k_B is the Boltzmann constant, and $\mathcal{W}(\xi)$ is the PMF along ξ defined by $\exp[-\beta \mathcal{W}(\xi)] \equiv \langle \delta(\xi(\mathbf{R}) - \xi) \rangle$.

During a WEUSMD simulation, an exchange between (neighboring) windows is regularly attempted. The exchange between windows i and j under window potentials w_m and w_n is accepted or rejected by the Metropolis criterion, $P_{ij} = \min[1, \exp(-\Delta)]$ where $\Delta = \beta\{[w_m(\xi_j) + w_n(\xi_i)] - [w_m(\xi_i) + w_n(\xi_j)]\}$. The average acceptance probability (P_a) of the exchange can be written as $P_a \equiv \iint d\xi_i d\xi_j p_m(\xi_i) p_n(\xi_j) P_{ij}$. One can easily find that the contributions to P_a from the guaranteed ($\Delta \leq 0$) and the conditional ($\Delta > 0$) exchanges are formally identical. Thus, in terms of Eq. (1), P_a can be written as

$$P_a = 2 \iint_{\Delta \leq 0} d\xi_i d\xi_j p_m(\xi_i) p_n(\xi_j) = 2 \frac{\iint_{\Delta \leq 0} d\xi_i d\xi_j e^{-\beta[\mathcal{W}(\xi_i) + \mathcal{W}(\xi_j) + w_m + w_n]}}{\iint d\xi_i d\xi_j e^{-\beta[\mathcal{W}(\xi_i) + \mathcal{W}(\xi_j) + w_m + w_n]}}. \quad (2)$$

Here, we omitted ξ_i in $w_m(\xi_i)$ [ξ_j in $w_n(\xi_j)$] for simplicity.

Assuming that $[\mathcal{W}(\xi_i) + \mathcal{W}(\xi_j)]$ is a slowly varying function compared to the window potentials in the region where the majority of ξ_i and ξ_j are populated, the PMF terms in the numerator and the denominator of Eq. (2) can be factored out. Since w is a harmonic function, the resulting equation becomes an integral of Gaussian functions

$$P_a = \frac{\sqrt{k_m k_n}}{\pi k_B T} \iint_{\Delta \leq 0} d\xi_i d\xi_j e^{-\beta[w_m(\xi_i) + w_n(\xi_j)]}. \quad (3)$$

The above approximation is reasonable because we consider the exchange between neighboring windows, and the contribution of integrands in Eq. (2) becomes negligible in the region where the window potential is sufficiently high. When $k_m = k_n = k$, which is the usual case for the umbrella sampling simulations, an analytic expression for Eq. (3) can be obtained as

$$P_a(z) = \mathrm{erfc}(z) + e^{-z^2} \sum_{j=0}^{\infty} \frac{{}_2F_1(1, 1/2, j+2, -1)}{\sqrt{2\pi}(j+1)!2^j} z^{2j+1}, \quad (4)$$

$$z = d\sqrt{\frac{k}{2k_B T}}, \quad (5)$$

where $d = |\xi_n - \xi_m|$, $\mathrm{erfc}(z) \equiv 1 - \mathrm{erf}(z)$ is the complementary error function, and

$${}_2F_1(a, b, c, d) = \sum_{l=0}^{\infty} \frac{(a)_l (b)_l}{(c)_l} \frac{d^l}{l!}, \quad (6)$$

with $(a)_l = a(a+1)\cdots(a+l-1)$. As shown in Fig. 1, $P_a(z)$ obtained using Eq. (4) agrees well with the P_a calculated from WEUSMD simulations of a TM helix-dimer model (see below). Indeed, Eq. (4) works well for various systems with different k and d (data not shown), indicating its general applicability.

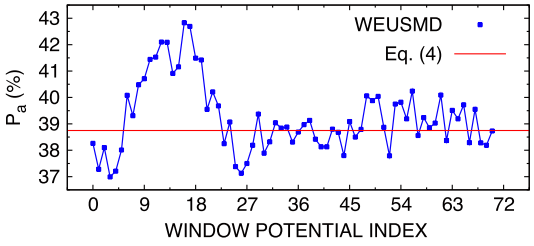

FIG. 1 (color online). The average acceptance probability $P_a(z)$ [Eq. (4)] compared with P_a calculated from the WEUSMD simulation for a pVNVV TM helix-dimer model.

To estimate the parameter set in an optimal range for WEUSMD, we adopt the first passage time optimization method [13], in which the optimized parameter set is determined by minimizing the mean round-trip time (τ_R) of a replica across temperature space [14,15]. Let us define the indices of N windows as $0, 1, \ldots, N-1$ ($d = \xi_{m+1} - \xi_m > 0$). Then, one can define the even pairs as $(0,1), (2,3), \ldots$ and the odd pairs as $(1,2), (3,4), \ldots$. At each step, the exchange pairs can be chosen either alternatively between the even and odd sequential pairs or randomly. In this study, only the former exchange-pair selection scheme is considered because it shows better performance for temperature REXMD (TREXMD) [14]. The τ_R of a window along ξ is minimized when

$$\frac{\partial \tau_R}{\partial z} \propto \frac{\partial}{\partial z}\left(\frac{1}{z^2}\left[\frac{1}{P_a(z)} - 1\right]\right) = 0. \quad (7)$$

The solution of Eq. (7) is $z_{\mathrm{opt}} = 0.8643$, and the corresponding P_a is 0.3875. With z_{opt}, the parameter set for WEUSMD (i.e., the relationship between d and k) is readily available from Eq. (5).

Using the optimal parameter set, we have applied WEUSMD to a pVNVV TM helix-dimer model [16] under the helix-helix distance restraint potential [17] and compared various results with those obtained from the TREXMD and USMD simulations to illustrate the efficacy and advantage of WEUSMD in TM helix assembly. The sequence of pVNVV is LLLLV LLLLL LNLLL LLLVL LLLLL VL, which is a membrane-soluble analogue of GCN4 leucine zipper (PDB:2ZTA) [18]. The helix-dimer interface involves hydrogen bonds (H bonds) between Asn residues [19,20]. The helix-helix distance (r_{HH}) and crossing angle (Ω) of the pVNVV dimer generated based on the leucine zipper are 9.23 Å and 29.3°, respectively [21]. To get the reference data, a 100-ns TREXMD simulation was performed with 16 replicas in a temperature range of 300–550 K starting from the configuration whose $r_{\mathrm{HH}} = 20$ Å and $\Omega = 29.3°$ (left-handed dimer). The initial configurations for WEUSMD and USMD were generated by translating each helix along r_{HH} and then rotating one helix randomly along its helical axis in order to randomize the TM helix contact interface. A total of 72 windows whose

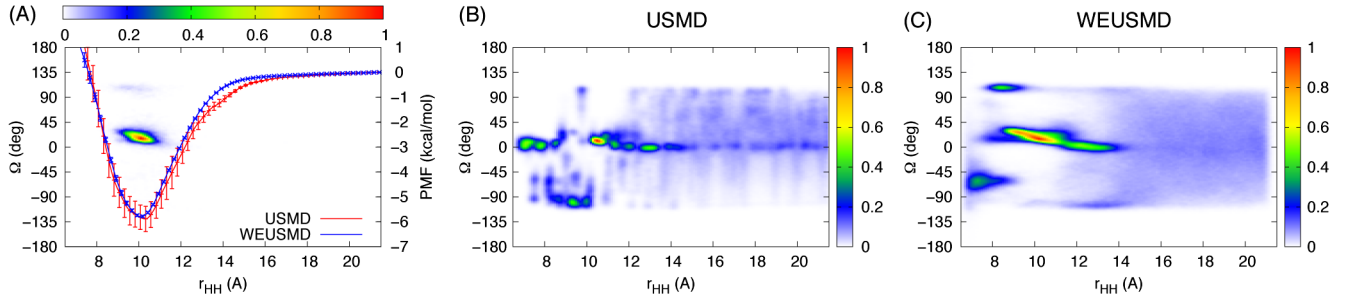

FIG. 2 (color online). (a) The population of r_{HH} and Ω, $P(r_{HH}, \Omega)$, sampled in TREXMD (density map) and the PMF as a function of r_{HH} from WEUSMD and USMD (solid lines). $P(r_{HH}, \Omega)$ was calculated using the ensemble structures at 300 K with the bin widths $\Delta r_{HH} = 0.1$ Å and $\Delta\Omega = 2°$, which are normalized by the highest population. The error bars for the PMFs are the standard deviations, which were calculated from the block averages of seven 10-ns trajectories. The population $P(r_{HH}, \Omega)$ from (b) USMD and (c) WEUSMD were calculated over all the windows with the same bin widths as those for TREXMD.

$\Omega = -29.3°$ (right-handed dimer) were generated in a r_{HH} range of 6.8–21 Å with 0.2 Å interval. With $d = 0.2$ Å and $z_{opt} = 0.8643$, an optimal force constant of $k = 22$ kcal/(mol · Å^2) was determined from Eq. (5) for the helix-helix distance restraint potential to restrain r_{HH} around each target value [17]. In this study, for computational efficiency, the EEF1/IMM1 implicit membrane model with a hydrophobic thickness of 23 Å was used to mimic a dimyristoylphosphatidylcholine (DMPC) membrane bilayer [22]. For each system, 100-ns Langevin dynamics simulations were performed. All simulations were performed using CHARMM [23] with the default IMM1 option, and window exchanges were controlled by the MMTSB toolset [24]. A time step of 2 fs was used for all the simulations with the SHAKE algorithm. The analysis of TREXMD was done for the last 90-ns trajectory, and those of WEUSMD and USMD were performed for the last 70-ns trajectory.

The population of the configuration space sampled at 300 K from TREXMD is shown as a function of r_{HH} and Ω, i.e., $P(r_{HH}, \Omega)$, in Fig. 2(a), and is bounded by $[9\text{ Å}, 11\text{ Å}] \times [0°, 40°]$ with a peak around $r_{HH} = 10.1$ Å and $\Omega = 16.4°$. In the same panel, the PMFs as a function of r_{HH} from WEUSMD and USMD are shown, which were calculated by the integration of the mean force acting along r_{HH} [16]. Each PMF minimum agrees well with the peak of $P(r_{HH}, \Omega)$, and the association free energy ΔG is -5.83 kcal/mol (WEUSMD) and -5.89 kcal/mol (USMD) in the IMM1 model. The thermally accessible range of r_{HH}, [9 Å, 11 Å], sampled in TREXMD, corresponds to the range where $|\mathcal{W}(r_{HH}) - \Delta G| \leq 2k_BT$ for WEUSMD and USMD. While both PMFs appear to be similar, the error bars up to $r_{HH} = 14$ Å are significantly larger in the USMD PMF than the WEUSMD PMF, indicating that the conformational and configurational sampling at short or intermediate r_{HH} is more efficient in WEUSMD and thus the PMF converges faster. Indeed, the 10-ns block average WEUSMD PMFs converged to the average PMF within 0.48 kcal/mol after the first 10-ns simulation, while the block average USMD PMFs showed a deviation up to 1.41 kcal/mol even after 100-ns simulations. In addition, a small deviation between the WEUSMD and USMD PMFs around $r_{HH} = 14$ Å arises from the incomplete sampling of USMD (see below).

The sampling efficiency in WEUSMD is illustrated in Figs. 2(b) and 2(c). In Fig. 2(b), USMD samples wide helix-helix configurations at $r_{HH} > 14$ Å where two TM helices do not strongly interact. But the configurational sampling in USMD is very restricted and strongly dependent on the initial configurations at $r_{HH} < 14$ Å (notably in the thermally accessible r_{HH}) due to the strong interactions between interfacial residues. As shown in Fig. 2(c), however, the configurations in the thermally accessible r_{HH} from WEUSMD agree well with those from TREXMD.

The enhanced sampling efficiency in WEUSMD arises from the fact that the configurations at different r_{HH} can be exchanged, which helps the system at each window overcome certain hidden (artificial) barriers introduced by the restraint potential and sample energetically favorable configurations. Therefore, other degrees of freedom for helix motion (Ω, helix rotation angles, etc.) can be sampled more efficiently by regular exchanges between windows along ξ (r_{HH} in this study). This feature is particularly important in TM helix assembly study for finding both critical TM-TM interfacial residues and the association energetics. In the case of the pVNVV dimer, since the Asn residues at the bilayer center can form H bonds, it is critical to examine how different sampling approaches describe such important interactions. Since the Asn residues show the closest contact in the thermally accessible r_{HH}, we simply consider the Asn side chain rotation angles θ_1 and θ_2 [see Fig. 3(a) for definition]. The population of θ_1 and θ_2, $P(\theta_1, \theta_2)$, from TREXMD in Fig. 3(b) shows three distinct peaks around $(\theta_1, \theta_2) = (-30°, 30°)$, $(-30°, -30°)$, and $(30°, -30°)$ with a minor peak around $(30°, 30°)$. As shown in Fig. 3(c), $P(\theta_1, \theta_2)$ sampled in WEUSMD agrees well with the TREXMD result. However, $P(\theta_1, \theta_2)$ sampled in USMD is significantly different [Fig. 3(d)]. These results clearly demonstrate

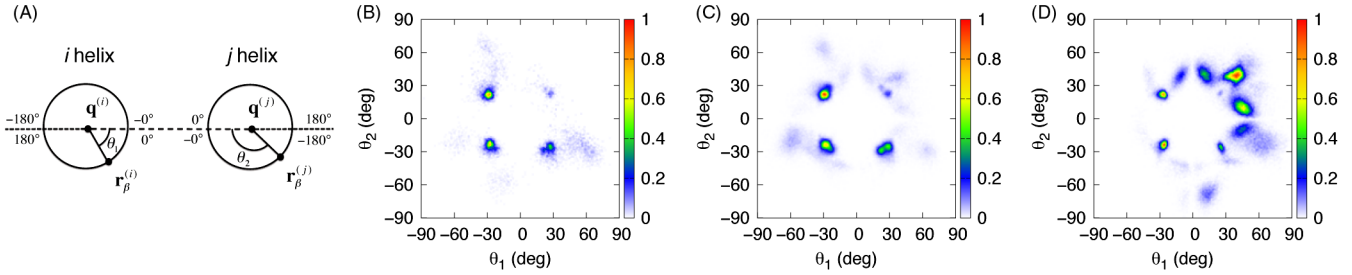

FIG. 3 (color online). (a) Asn side chain rotational angles θ_1 and θ_2. For two helices i and j, θ_1 is defined by the angle between $\mathbf{q}^{(j)} - \mathbf{q}^{(i)}$ and $\mathbf{r}_\beta^{(i)} - \mathbf{q}^{(i)}$, where $\mathbf{q}^{(i)}$ and $\mathbf{q}^{(j)}$ are the projections of $\mathbf{r}_\beta^{(i)}$, the Asn C_β position in helix i, and $\mathbf{r}_\beta^{(j)}$ onto their helix axes, respectively. θ_2 is defined similarly. The population of θ_1 and θ_2, $P(\theta_1, \theta_2)$, sampled from (b) TREXMD, (c) WEUSMD, and (d) USMD. In the calculation of $P(\theta_1, \theta_2)$ for WEUSMD and USMD, only the configurations with r_{HH} in the thermally accessible range, [9 Å, 11 Å], were considered.

that WEUSMD can be an efficient method in finding optimal interfaces for TM helix assembly (in a given potential energy function) due to the facilitated sampling of unrestrained degrees of freedom for helix motion.

In summary, the application of WEUSMD to a pVNVV TM dimer model demonstrates its efficacy and advantage in searching critical interfaces for TM helix assembly (in a given potential energy function) without sacrificing the availability of the PMF along the reaction coordinate for helix motion. Starting from an initially randomized TM-TM orientation, WEUSMD was able to sample very similar H-bonding patterns as in TREXMD, in addition to the PMF along the helix-helix distance. Although the PMF calculated from USMD is comparable to that from WEUSMD, USMD was not able to explore the configuration space for optimal interfaces of helix assembly due to the strong interactions between residues at short and intermediate separations. As the terminology implies, WEUSMD combines the advantages of REXMD and USMD. A notable advantage of WEUSMD is that it does not require extra simulations, which is typical in TREXMD (replicas at different T than that of interest). In addition, with a predetermined parameter set in an optimal range by Eqs. (5) and (7), the iterative determination of an optimal parameter set can be avoided, which makes its application simple and straightforward. We have developed various helix restraint potentials [25,26]; therefore, it is possible to use any of these restraints or their combinations in WEUSMD for efficient and important sampling, especially in all-atom explicit membranes, to dissect the helix-helix and helix-lipid contributions along various helix motions in lipid membranes.

This work was supported by NIH R01-GM092950 and TeraGrid resources provided by Purdue University (NSF OCI-0503992).

*wonpil@ku.edu

[1] M. S. Almén, K. J. V. Nordström, R. Fredriksson, and H. B. Schiöth, BMC Biol. **7**, 50 (2009).
[2] R. M. Bill, P. J. F. Henderson, S. Iwata, E. R. S. Kunji, H. Michel, R. Neutze, S. Newstead, B. Poolman, C. G. Tate, and H. Vogel, Nat. Biotechnol. **29**, 335 (2011).
[3] D. Dell'Orco, P. G. De Benedetti, and F. Fanelli, J. Phys. Chem. B **111**, 9114 (2007).
[4] J. Hènin, A. Pohorille, and C. Chipot, J. Am. Chem. Soc. **127**, 8478 (2005).
[5] R. H. Swendsen and J.-S. Wang, Phys. Rev. Lett. **57**, 2607 (1986).
[6] K. Hukushima and K. Nemoto, J. Phys. Soc. Jpn. **65**, 1604 (1996).
[7] U. H. E. Hansmann, Chem. Phys. Lett. **281**, 140 (1997).
[8] Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **314**, 141 (1999).
[9] D. M. Zuckerman, Annu. Rev. Biophys. **40**, 41 (2011).
[10] F. G. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050 (2001).
[11] L. Zheng, M. Chen, and W. Yang, Proc. Natl. Acad. Sci. U.S.A. **105**, 20227 (2008).
[12] Y. Sugita, A. Kitao, and Y. Okamoto, J. Chem. Phys. **113**, 6042 (2000).
[13] W. Nadler and U. H. E. Hansmann, Phys. Rev. E **75**, 026109 (2007).
[14] M. Lingenheil, R. Denschlag, G. Mathias, and P. Tavan, Chem. Phys. Lett. **478**, 80 (2009).
[15] W. Nadler and U. H. E. Hansmann, J. Phys. Chem. B **112**, 10 386 (2008).
[16] J. Lee and W. Im, J. Am. Chem. Soc. **130**, 6456 (2008).
[17] J. Lee and W. Im, J. Comput. Chem. **28**, 669 (2007).
[18] E. K. O'shea, J. Klemn, P. Kim, and T. Alber, Science **254**, 539 (1991).
[19] C. Choma, H. Gratkowski, J. D. Lear, and W. F. DeGrado, Nat. Struct. Biol. **7**, 161 (2000).
[20] F. X. Zhou, M. J. Cocco, W. P. Russ, A. T. Brunger, and D. M. Engelman, Nat. Struct. Biol. **7**, 154 (2000).
[21] C. Chothia, M. Levitt, and D. Richardson, J. Mol. Biol. **145**, 215 (1981).
[22] T. Lazaridis, Proteins **52**, 176 (2003).
[23] B. R. Brooks *et al.*, J. Comput. Chem. **30**, 1545 (2009).
[24] M. Feig, J. Karanicolas, and C. L. Brooks III, J. Mol. Graphics Modell. **22**, 377 (2004).
[25] J. Lee and W. Im, Phys. Rev. Lett. **100**, 018103 (2008).
[26] W. Im, J. Lee, T. Kim, and H. Rui, J. Comput. Chem. **30**, 1622 (2009).